



University of Groningen

Metagenomic Characterization of the Human Intestinal Microbiota in Fecal Samples from STEC-Infected Patients

Gigliucci, Federica; von Meijenfeldt, F. A. Bastiaan; Knijn, Arnold; Michelacci, Valeria; Scavia, Gaia; Minelli, Fabio; Dutilh, Bas E.; Ahmad, Hamideh M.; Raangs, Gerwin C.; Friedrich, Alex W.

Published in:
Frontiers in Cellular and Infection Microbiology

DOI:
[10.3389/fcimb.2018.00025](https://doi.org/10.3389/fcimb.2018.00025)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Gigliucci, F., von Meijenfeldt, F. A. B., Knijn, A., Michelacci, V., Scavia, G., Minelli, F., ... Morabito, S. (2018). Metagenomic Characterization of the Human Intestinal Microbiota in Fecal Samples from STEC-Infected Patients. *Frontiers in Cellular and Infection Microbiology*, 8, [25].
<https://doi.org/10.3389/fcimb.2018.00025>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Metagenomic Characterization of the Human Intestinal Microbiota in Fecal Samples from STEC-Infected Patients

Federica Gigliucci^{1,2*}, F. A. Bastiaan von Meijenfeldt³, Arnold Knijn¹, Valeria Michelacci¹, Gaia Scavia¹, Fabio Minelli¹, Bas E. Dutilh^{3,4}, Hamideh M. Ahmad⁵, Gerwin C. Raangs⁵, Alex W. Friedrich⁵, John W. A. Rossen⁵ and Stefano Morabito¹

¹ Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanità, Viale Regina Elena, Rome, Italy, ² Department of Sciences, University Roma Tre, Rome, Italy, ³ Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, Netherlands, ⁴ Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, Nijmegen, Netherlands, ⁵ Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

OPEN ACCESS

Edited by:

Alfredo G. Torres,
University of Texas Medical Branch,
United States

Reviewed by:

Maite Muniesa,
University of Barcelona, Spain
Analia Inés Etcheverría,
National University of Central Buenos
Aires, Argentina
Wessam Galia,
UMR5557 Ecologie Microbienne
(LEM), France

*Correspondence:

Federica Gigliucci
federica.gigliucci@libero.it

Received: 30 October 2017

Accepted: 18 January 2018

Published: 06 February 2018

Citation:

Gigliucci F, von Meijenfeldt FAB, Knijn A, Michelacci V, Scavia G, Minelli F, Dutilh BE, Ahmad HM, Raangs GC, Friedrich AW, Rossen JWA and Morabito S (2018) Metagenomic Characterization of the Human Intestinal Microbiota in Fecal Samples from STEC-Infected Patients. *Front. Cell. Infect. Microbiol.* 8:25. doi: 10.3389/fcimb.2018.00025

The human intestinal microbiota is a homeostatic ecosystem with a remarkable impact on human health and the disruption of this equilibrium leads to an increased susceptibility to infection by numerous pathogens. In this study, we used shotgun metagenomic sequencing and two different bioinformatic approaches, based on mapping of the reads onto databases and on the reconstruction of putative draft genomes, to investigate possible changes in the composition of the intestinal microbiota in samples from patients with Shiga Toxin-producing *E. coli* (STEC) infection compared to healthy and healed controls, collected during an outbreak caused by a STEC O26:H11 infection. Both the bioinformatic procedures used, produced similar result with a good resolution of the taxonomic profiles of the specimens. The stool samples collected from the STEC infected patients showed a lower abundance of the members of *Bifidobacteriales* and *Clostridiales* orders in comparison to controls where those microorganisms predominated. These differences seemed to correlate with the STEC infection although a flexion in the relative abundance of the *Bifidobacterium* genus, part of the *Bifidobacteriales* order, was observed also in samples from Crohn's disease patients, displaying a STEC-unrelated dysbiosis. The metagenomics also allowed to identify in the STEC positive samples, all the virulence traits present in the genomes of the STEC O26 that caused the outbreak as assessed through isolation of the epidemic strain and whole genome sequencing. The results shown represent a first evidence of the changes occurring in the intestinal microbiota of children in the course of STEC infection and indicate that metagenomics may be a promising tool for the culture-independent clinical diagnosis of the infection.

Keywords: STEC, microbiota, human gut, HUS, diarrhea, metagenomics

Abbreviations: STEC, Shiga toxin-producing *Escherichia coli*; HUS, Haemolytic Uremic Syndrome; CD, Crohn's Disease; EPEC, Enteropathogenic *Escherichia coli*; OTUs, Operational Taxonomic Units.

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are a heterogeneous *E. coli* pathogroup causing food-borne outbreaks and sporadic cases of disease worldwide (Armstrong et al., 1996). STEC may cause severe afflictions in humans due to their ability to produce potent cytotoxins, the Shiga toxins (Stx), acquired upon infection with bacteriophages carrying stx genes, which can remain stably integrated into the bacterial chromosome (O'Brien et al., 1984). Stx exert their action by blocking the protein synthesis in the target cells by inactivating ribosomes (Okuda et al., 2006). Upon infection, the host can present a wide range of symptoms, including uncomplicated diarrhea, haemorrhagic colitis and the life-threatening haemolytic uremic syndrome (HUS).

The pathogenesis of STEC infections is not completely understood as it seems that, beside the virulence potential of the infecting strains, a number of other factors appear to be involved in the progression of the clinical symptoms. One possibility is that the human intestinal microbiota play a role by interfering with the ability of STEC to efficiently colonize the gastro-intestinal tract, as it has been proposed for other bacterial infections (Fujiwara et al., 2001; Gueimonde et al., 2007). Additionally, Gamage and colleagues proposed that different bacterial species in the host microbiota can act as amplifiers of the Stx-converting phage resulting in an augmented ability to produce the toxin (Gamage et al., 2003).

More and more data are becoming available on the role of human microbiota in health and disease and there is increasing evidence that the commensal bacteria play a crucial role in protecting human health (Forbes et al., 2016). Indeed, the human gut microbiota is a homeostatic ecosystem with several vital functions essential to host health, including protection against pathogens (Shreiner et al., 2015). Tap and colleagues showed that the human gut microbiota is governed by the presence of *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, *Proteobacteria* and, in some cases *Verrucomicrobia* bacterial phyla (Tap et al., 2009). Members of these taxa, including *Faecalibacterium*, *Ruminococcus*, *Eubacterium*, *Dorea*, *Bacteroides*, *Alistipes*, and *Bifidobacterium* genera, constitute a phylogenetic core shared among individuals (Tap et al., 2009). In particular, it has been shown that species belonging to *Bifidobacterium* genus and butyrate-producing bacteria, belonging to the *Clostridiales* order, might exert a variety of beneficial health effects (O'Callaghan and van Sinderen, 2016; Rivi re et al., 2016). Hence, a decrease in the relative abundances of *Bifidobacterium* species in the human colon has been associated with several disorders, such as inflammatory bowel disease, Crohn's disease and ulcerative colitis, irritable bowel syndrome, colorectal cancer, and increased gut permeability (O'Callaghan and van Sinderen, 2016). The mentioned disorders lead to a general change in the gut microbiota composition, favoring also the colonization and proliferation of pathogenic microorganisms (de Vos and de Vos, 2012).

We used a shotgun metagenomic sequencing approach to investigate the taxonomic composition of the gut microbiota in fecal samples taken from patients suffering from STEC infection

and compared the results with those found in samples from healthy controls. Additionally, we analyzed fecal samples from patients with Crohn's disease with and without evidence of infection with STEC and added the microbial profiles of these samples to the analysis in order to assess if any change observed in the composition of the gut microbiota in course of STEC infections may be related with a dysbiotic status rather than associated with the STEC infection itself.

MATERIALS AND METHODS

Samples Origin

Fecal samples ($N = 10$) from children (ages 0–4) were collected during an outbreak of STEC infections in 2015 in a nursery in the province of Rome, Italy. During the investigation, a STEC O26:H11 was isolated from different patients. Three samples were from patients with diarrhea (A.9, A.8, A.14), three were collected from patients after 2 weeks from the restoration of the normal intestinal function (A.40, A.32, A.41), and four samples (A.4, A.30, A.16, 481-5) from healthy subjects. In addition, four stool samples from four patients (ages 10–18) suffering from Crohn's Disease (CD) and hospitalized at the University Medical Center Groningen, The Netherlands, were included in the analysis (Samples 1, 2, 5, and 6). These samples were previously analyzed by real-time PCR for the presence of virulence genes associated with STEC (*stx1* and *stx2*) and EPEC (*escV*) (Gauthier et al., 2003) pathogroups, revealing the presence of *stx2* gene in Sample 2, the presence of the *escV* gene in Sample 1 and 5, and none of the mentioned virulence genes in Sample 6. The CD samples used for the present analyses were collected in the course of routine diagnostics and infection prevention controls. Oral consent for the use of such clinical samples for research purposes is routinely obtained upon patient admission to the UMCG, in accordance with the guidelines of the Medical Ethics Committee of the University Medical Center Groningen. The experiments, accordingly with the guidelines of the Declaration of Helsinki and the institutional regulations, were performed on anonymized samples.

DNA Extraction and Sequencing

DNA of the specimens related with the Italian outbreak was extracted from 0.20 g of each stool sample using the EZNA Stool DNA extraction kit (Omega Bio-tek, Norcross, Ga.) following manufacturer's instructions, whereas for the CD samples, DNA was extracted from 0.25 g feces using the Power Soil DNA isolation kit (MO BIO Laboratories inc., Carlsbad, CA, USA) following the manufacturer's instructions. We did not observe marked differences in the purity of DNA produced with the two methods as assessed by considering the ratio between the absorbance measured at 260 and 280 nm.

Sequencing libraries were prepared from 100 ng of the DNA extracted from samples **A.9, A.4, and A.30**, using the NEBNext Fast DNA Fragmentation & Library Prep kit (New England BioLabs, New England, USA). In detail, the DNA was enzymatically fragmented to obtain fragments of about 400 bp, through an incubation at 25°C for 15 min, followed by 10 min at 70°C. The fragmented DNA was subjected to link with adapters

and size selection of 450 bp fragments by electrophoresis on E-Gel SizeSelect 2% (Invitrogen, Carlsbad, USA) followed by PCR amplification as indicated in the NEBNext Fast DNA Fragmentation & Library Prep kit manual (New England BioLabs, USA). The libraries were amplified individually through emulsion PCR with an Ion OneTouch 2 and sequenced with an Ion Torrent Personal Genome Machine (Life Technologies, 118 Carlsbad, USA), using the 400 bp sequencing protocol. The three samples were sequenced individually in three different runs using a 316 V2 chip per run.

Sequencing of the samples **A.8, A.14, A.40, A.32, A.41, A.16, 481-5** as well as of the four CD fecal samples DNA was carried out using the TruSeq Nano DNA Library Preparation kit (Illumina, San Diego, CA, US) and a MiSeq platform (Illumina, San Diego, CA, US). In detail, sequencing libraries were prepared from 100 ng of the DNA extracted from each sample. DNA was mechanically fragmented to obtain a 350 bp insert size, using the M220 Focused-ultrasonicator™ based on Covaris AFA (Adaptive Focused Acoustics™) technology. The fragmented DNA was subjected to end repair and size selection of 350 bp fragments, followed by adenylation of 3' ends, link with adaptors and a final enrichment of DNA fragments, following the TruSeq Nano DNA Library Preparation kit manual (Illumina, San Diego, CA, US). The 11 samples were sequenced in three different runs, using a 600V3 cartridge per run generating 300 bp paired-end reads.

All the metagenomes are available at European Nucleotide Archive at EMBL-EBI under the accession number PRJEB23207.

Bioinformatic Analyses

Read Mapping Analysis

Reference-based bioinformatic analyses of the metagenomes were performed using the tools available on the ARIES public webserver (<https://www.iss.it/site/aries/>).

In detail, raw sequence reads were subjected to a quality check and trimmed to remove the adaptors and to accept 20 as the lowest Phred value. The identification of the presence of *E. coli* virulence genes was performed through the pipeline Virulotyper, which employs the Bowtie2 v2.3.2.2 program (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) (Langmead and Salzberg, 2012) to map the sequencing reads against the *E. coli* Virulence genes database (Joensen et al., 2014). Virulence genes showing average coverage above 1X were considered to be present in the sample.

Taxonomic analysis was performed using the DIAMOND v0.8.24 tool (Buchfink et al., 2015) to align the reads in FASTA format to the NCBI-nr (non-redundant) database (<ftp://ftp.ncbi.nlm.nih.gov/>) downloaded on July 2016, with the DIAMOND-BLASTX algorithm. Visualization was done with MEGAN (Meta Genome ANalyzer) version 6 (<http://www-ab.informatik.uni-tuebingen.de/software/megan6/>) (Huson et al., 2007).

The operational taxonomic units (OTUs) content of the samples was determined using MEGAN 6 and the converter script prot-gi2taxid-August2016X.bin downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov>). MEGAN 6 was also used to perform rarefaction analysis (Gotelli and Colwell, 2001).

Reference Free Analysis

The same metagenomes were also analyzed using a novel approach, as described by Kang et al. (2015), which allows assessing the degree of similarity between complex microbial communities through the reconstruction of draft genome sequences from shotgun metagenomic sequencing. Reads from the different metagenomes obtained with the Illumina platform were de novo assembled. The assembled genomic fragments were eventually grouped in putative genomes (bins) using probabilistic distances. For a visualization of the binning process we refer to Figure 1 in Kang et al. (2015). Default settings were used for all programs, unless otherwise mentioned.

In detail, the paired-end trimmed Illumina reads were cross-assembled with SPAdes v3.10.1 (<http://bioinf.spbau.ru/spades>) (Bankevich et al., 2012) in “—meta” mode (Nurk et al., 2017) and used as reference for mapping of all the single metagenomes, including those obtained with the Ion Torrent PGM. The quality of the cross-assembled scaffolds was evaluated using the QQuality ASsessment Tool (QUAST) v4.5 (<http://quast.sourceforge.net/>) (Gurevich et al., 2013). QUAST provides assembly statistics including the number of assembled contigs, the length of the longest contig and the values of N50, N75, L50, L75, GC (%).

The reads of the metagenomes were mapped to the cross-assembled scaffolds using Burrows-Wheeler Aligner (BWA) v0.7.15, with the BWA-MEM algorithm (Li and Durbin, 2010). BWA-MEM was run with default settings, thus distributing reads that map to multiple places evenly, as suggested in the MetaBAT usage manual (<https://bitbucket.org/berkeleylab/metabat>). The output of BWA-MEM was converted from SAM format to the binary BAM format with the tools provided in the SAMtools suite (Li et al., 2009). The percentage of the reads mapping against the cross-assembly was estimated for the different metagenomes using SAMtools flagstat. The open source software MetaBAT2 (Metagenome Binning with Abundance and Tetra-nucleotide frequencies) v2.9.1 (<http://bitbucket.org/berkeleylab/metabat>) (Kang et al., 2015) was used to obtain bins from the scaffolds. MetaBAT2 was fed with the cross-assembly as input together with a depth file based on the bam files. The depth file was generated with the script `jgi_summarize_bam_contig_depths`, that is supplied with MetaBAT. The quality of the genomic bins generated was assessed with CheckM v1.0.5 (Parks et al., 2015) in “—lineage_wf” mode. CheckM assesses bin completeness, contamination, and strain heterogeneity based on the absence and presence of sets of expected single copy marker genes. Bins with a completeness >35% and a contamination <5% were selected for further analyses.

The scaffolds were annotated with the CAT (The Contig Annotation Tool) pipeline (Cambuy et al., 2016) and used to taxonomically classify the bins. To account for possible conflicts within a bin, we only annotated a bin to a taxonomic level if at least half of the length of the sequences within the bin showed a consistent annotation. If no annotation reached majority, we annotated the bin to the annotation with the longest length representation, but marked that annotation as “possible.” For instance, if 70% of the length of the bin is annotated on the genus level to *Escherichia*,

but only 30% is annotated to *E. coli* on the species level, that bin is annotated as an *Escherichia* bacterium, possibly *E. coli*.

To calculate relative abundance of the bins in the samples, we used the depth file generated earlier with `jgi_summarize_bam_contig_depths`. Average coverage of a bin in a sample was calculated by multiplying the depth of the scaffolds in that bin by their length, and dividing their sum by the total base pair length of all the scaffolds in the bin. Abundance was made relative per sample by dividing this average read coverage per base pair by the sum of average read coverage per base pair for all the bins. Bin abundance was normalized to account for the fact that not all reads mapped to bins, by multiplying relative abundance with the fraction of binned reads, which was calculated as the sum of reads mapping to binned scaffolds (as calculated with `samtools idxstats`) divided by the total number of reads mapping to the cross-assembly (as calculated with `samtools flagstat`). Thus, the sum of relative abundances of all bins adds up to the fraction of total reads mapping to the cross-assembly that map to the bins in a sample.

The scaffolds were searched for the occurrence of the genes present in the *E. coli* Virulence genes database (Joensen et al., 2014). The database was queried with `tblastx v2.6.0+` (Camacho et al., 2009) against the scaffolds. A gene was considered present

on a scaffold if its hit had an e-value below $1e-5$ and query coverage of at least 70%.

RESULTS

Detection of Virulence Genes Associated with Pathogenic *E. coli* in the Metagenomes

Detection of the STEC O26:H11 Virulence Genes

The virulence features of the STEC O26:H11 strain that had caused the outbreak had been previously identified by characterizing the isolated strain through real-time PCR and Whole Genome Sequencing (unpublished). In this work, we first checked the capability of metagenomics to identify the presence of the genes associated with the epidemic STEC strain. Mapping of the quality-controlled reads against the *E. coli* Virulence genes database (Joensen et al., 2014) showed the presence of all the genes composing the virulome of the STEC O26 outbreak strain (Table 1). In one sample (A.9), the metagenomics did not show the presence of the *espF* gene, while in sample A.14 we identified the presence of the gene *iroN*, which however was not present in the outbreak strain (Table 1). This gene was later identified in a scaffold that was annotated to the family *Enterobacteriaceae*, but was unbinned.

TABLE 1 | Comparison of the STEC virulence genes identified in the samples collected from patients with STEC infection by metagenomics with those obtained through WGS of the isolated outbreak strain.

| Metagenomic analysis | | | | | | | | | | Isolation based method |
|------------------------|--------------------|--------------------|--------------------|-------------|--------------------|--------------------|-------------|-------|-------------|------------------------|
| Patients with diarrhea | | | Recovered patients | | | Healthy subjects | | | | Epidemic strain |
| A. 9 | A. 8 | A. 14 | A. 40 | A. 41 | A. 32 | A. 4 | A. 16 | 481-5 | A. 30 | |
| <i>Cif</i> | <i>cif</i> | <i>cif</i> | <i>ehxA</i> | <i>mchF</i> | <i>katP</i> | <i>epeA</i> | <i>cba</i> | | <i>senB</i> | <i>cif</i> |
| <i>Eae</i> | <i>eae</i> | <i>eae</i> | <i>espP</i> | <i>tsh</i> | <i>lpfA</i> | <i>espP</i> | <i>cma</i> | | | <i>eae</i> |
| <i>efa1</i> | <i>efa1</i> | <i>efa1</i> | <i>katP</i> | | | <i>gad</i> | <i>gad</i> | | | <i>efa1</i> |
| <i>ehxA</i> | <i>ehxA</i> | <i>ehxA</i> | <i>mchC</i> | | | <i>iha</i> | <i>ireA</i> | | | <i>ehxA</i> |
| <i>espA</i> | <i>espA</i> | <i>espA</i> | <i>pic</i> | | | | <i>iss</i> | | | <i>espA</i> |
| <i>espB</i> | <i>espB</i> | <i>espB</i> | <i>senB</i> | | | | <i>lpfA</i> | | | <i>espB</i> |
| <i>espJ</i> | <i>espF</i> | <i>espF</i> | <i>toxB</i> | | | | <i>pic</i> | | | <i>espF</i> |
| <i>espP</i> | <i>espJ</i> | <i>espJ</i> | | | | | <i>vat</i> | | | <i>espJ</i> |
| <i>gad</i> | <i>espP</i> | <i>espP</i> | | | | | | | | <i>espP</i> |
| <i>iha</i> | <i>gad</i> | <i>gad</i> | | | | | | | | <i>gad</i> |
| <i>iss</i> | <i>iha</i> | <i>iha</i> | | | | | | | | <i>iha</i> |
| <i>katP</i> | <i>iss</i> | <i>iroN</i> | | | | | | | | <i>iss</i> |
| <i>lpfA</i> | <i>katP</i> | <i>iss</i> | | | | | | | | <i>katP</i> |
| <i>nleA</i> | <i>lpfA</i> | <i>katP</i> | | | | | | | | <i>lpfA</i> |
| <i>nleB</i> | <i>nleA</i> | <i>lpfA</i> | | | | | | | | <i>nleA</i> |
| <i>nleC</i> | <i>nleB</i> | <i>nleA</i> | | | | | | | | <i>nleB</i> |
| <i>stx2a</i> | <i>nleC</i> | <i>nleB</i> | | | | | | | | <i>nleC</i> |
| <i>tir</i> | <i>stx2a</i> | <i>nleC</i> | | | | | | | | <i>stx2a</i> |
| <i>toxB</i> | <i>tir</i> | <i>stx2a</i> | | | | | | | | <i>tir</i> |
| | <i>toxB</i> | <i>tir</i> | | | | | | | | <i>toxB</i> |
| | | <i>toxB</i> | | | | | | | | |

Plasmidic genes normally present on the large virulence plasmid of STEC O26 are indicated in bold.

The genes encoding the Stx were not identified in the specimens collected from healthy subjects and from recovered patients. However, we could observe in some of the latter metagenomes the presence of genes carried by the large virulence plasmid of STEC and identified also in the outbreak strain (**Table 1**). In particular, the metagenome from sample A.40 displayed the presence of plasmidic genes *ehxA* (Beutin et al., 1990), *espP* (Brunder et al., 1997), *katP* (Caprioli et al., 2005), and *toxB* (Tatsuno et al., 2000), while that from sample A.32 had the gene *katP* (Caprioli et al., 2005) (**Table 1**). Finally, we could not observe the presence of any gene associated with STEC in any of the metagenomes from healthy subjects, with the exception of the presence of *espP* in sample A.4 and *lpfA* in sample A.16 (**Table 1**).

The same analysis carried out on the metagenomes of the samples collected from patients with Crohn's disease confirmed the previous real-time PCR results (**Table 2**). In detail, Sample 6 appeared negative for the presence of virulence genes associated with pathogenic *E. coli* infections; Sample 2 showed the presence of the Stx2f encoding genes, confirming the previous evidence of STEC infection; finally, the remaining two samples, although they did not show the presence of the *escV* gene, previously identified by real-time PCR, displayed a virulence genes profile compatible with the presence of an aEPEC strain (**Table 2**).

Detection of Virulence Genes in the Cross-Assembled Scaffolds

In a complementary bioinformatic analysis of the same data, we assembled the metagenomic reads into scaffolds, and binned the scaffolds into draft genome sequences (Garza and Dutilh, 2015). The cross-assembly of the Illumina metagenomes, including the samples from STEC infections and the related control group as well as those from Crohn's disease, produced 429,862 scaffolds of size ≥ 500 bp, where the longest scaffold was 505,337 bp (Table S1). The percentage of the reads mapping against the assembled scaffolds was comparable between the different metagenomes, and seemed only slightly influenced by sequencing platform, ranging from 85.96 to 99.77% (Table S2). It is important to note that the cross-assembly is only based on the Illumina sequences, but the Ion Torrent reads showed high mappability values as well (Table S2). Finally, the metagenome binning produced 209 draft genome bins in total.

The tblastx search confirmed the presence of *E. coli* virulence genes in the cross-assembled metagenomic scaffolds, that were already detected in the metagenomes based on mapping the quality-controlled reads on the virulence genes database as described above. Additionally, this analysis allowed to localize each gene in specific scaffolds and within binned draft genomes. Most of the assembled scaffolds that contained virulence genes

TABLE 2 | Comparison of virulence genes associated with pathogenic *E. coli* infection identified by metagenomics and real-time PCR in the samples collected from patients with Crohn's disease.

| Sample 1 | | Sample 2 | | | Sample 5 | | Sample 6 | |
|----------------------------|---------------|----------------------------|----------------------------|---------------|----------------------------|---------------|----------------------------|---------------|
| Metagenomics reads mapping | Real time PCR | Metagenomics reads mapping | Metagenomics bin alignment | Real time PCR | Metagenomics reads mapping | Real time PCR | Metagenomics reads mapping | Real time PCR |
| <i>cba</i> | <i>escV</i> | <i>cif</i> | <i>eae</i> | <i>stx2f</i> | <i>cif</i> | <i>escV</i> | | |
| <i>cif</i> | | <i>cnf1</i> | <i>gad</i> | | <i>espA</i> | | | |
| <i>cma</i> | | <i>eae</i> | <i>hlyE</i> | | <i>espF</i> | | | |
| <i>eae</i> | | <i>espA</i> | <i>iha</i> | | <i>gad</i> | | | |
| <i>espA</i> | | <i>espC</i> | <i>ireA</i> | | <i>nleB</i> | | | |
| <i>espJ</i> | | <i>espF</i> | <i>iroN</i> | | | | | |
| <i>espD</i> | | <i>espJ</i> | <i>katP</i> | | | | | |
| <i>gad</i> | | <i>fim41a</i> | <i>lpfA</i> | | | | | |
| <i>iroN</i> | | <i>gad</i> | <i>mchF</i> | | | | | |
| <i>iss</i> | | <i>iha</i> | <i>pet</i> | | | | | |
| <i>katp</i> | | <i>ireA</i> | <i>prfB</i> | | | | | |
| <i>lpfA</i> | | <i>iroN</i> | <i>sat</i> | | | | | |
| <i>mchF</i> | | <i>iss</i> | | | | | | |
| <i>pet</i> | | <i>lpfA</i> | | | | | | |
| <i>tir</i> | | <i>mchB</i> | | | | | | |
| <i>tsh</i> | | <i>mchC</i> | | | | | | |
| <i>vat</i> | | <i>mchF</i> | | | | | | |
| | | <i>mcmA</i> | | | | | | |
| | | <i>nleB</i> | | | | | | |
| | | <i>nleC</i> | | | | | | |
| | | <i>pic</i> | | | | | | |
| | | <i>stx2f</i> | | | | | | |
| | | <i>tir</i> | | | | | | |
| | | <i>vat</i> | | | | | | |

that were present in the epidemic strain were unbinned, but were annotated as *E. coli* on the scaffold level. Moreover, all of the plasmidic genes that were seen in recovered patients with read mapping were also unbinned. It is common that plasmids are not associated with draft genomes binned from metagenomes, as their genomic signals including abundance and nucleotide usage are different from those of the core genome (Beitel et al., 2014; Kang et al., 2015). Not all genes seen with read mapping were found in the scaffolds, the most notable omission being *ehxA*.

Taxonomic Profiling of the Microbiota from Stool Samples

The metagenomic sequencing produced datasets of different sizes, mainly due to the use of two different sequencing platforms. As for the samples from STEC infections and the related control group, an average of 3,563,158 reads per sample were obtained from the Ion Torrent sequencing, while an average of 12,287,432 reads per sample were obtained from the Illumina platform. Nevertheless, the rarefaction curves (Gotelli and Colwell, 2001) showed that the diversity of taxonomic units was comparable between the different samples (Figure S1). The metagenomic sequencing of the four stool samples from Crohn's disease patients, produced with the Illumina sequencer produced on average 16,429,446 reads per sample.

Taxonomic Profiling Based on Read Mapping

In the metagenomes assayed, the most abundant phyla of *Bacteria* were the *Firmicutes* (16.6–79%), *Bacteroidetes* (0.2–63.1%), *Proteobacteria* (1.65–56.6%), and *Actinobacteria* (0.8–56.2%), whereas the phylum *Verrucomicrobia* showed a high abundance only in one sample (27.4% in A.30), obtained from a healthy subject (Table 3). In the samples analyzed the *Bifidobacteriales* order of the *Actinobacteria* phylum appeared more abundant in the vast majority of the STEC-negative samples (Figure 1) and a deeper analysis showed a marked prevalence of *Bifidobacterium longum* species (Figure S2). In addition to *Actinobacteria*, also some orders of the phylum *Firmicutes* showed a different distribution in the two groups of samples (Figure 1) ($p < 0.05$). The *Clostridiales* (*Clostridia*) were more abundant in the control group, while the *Lactobacillales* (*Bacilli*) predominated in the cases group (Figure 1) ($p < 0.05$) with the exception of sample A.9, which showed a taxonomic profile more similar to that of the control group with lower *Lactobacillales* (Figure 1). Moreover, the *Roseburia*, *Coprococcus*, *Butyrivibrio*, and *Faecalibacterium*, previously shown to be common in the intestinal microbiota from healthy subjects (Rivière et al., 2016; Hugon et al., 2017) were the most abundant genera of *Clostridiales* in the STEC negative samples assayed in this study (Figure 2), except for sample A.9, which again showed a profile more similar to those obtained from the control group samples. Members of the *Proteobacteria* and *Bacteroidetes* phyla apparently were not concerned by the perturbation of the intestinal microflora following the STEC infection as their relative proportions did not show patterns attributable to specific groups in the metagenomes analyzed (Figure 1).

The most abundant bacterial taxa identified in the specimens collected from Crohn's disease patients were the *Bacteroidetes*

(10.5–84%), *Proteobacteria* (4–77%), *Firmicutes* (9–29.5%), and *Actinobacteria* (1–26%) (Figure 3A). In one sample, a marked prevalence of the *E. coli* species was observed (Sample 2 in Figure 3). Interestingly, this sample was positive for STEC-associated genes, both in real-time PCR and at the metagenomic analysis.

Analysis of Genome Bins

For the analysis of the samples from STEC infections and the related controls, we selected 18 of the 209 bins based on values of completeness $>35\%$, contamination $<5\%$, and abundance $>5\%$ in at least one sample.

The abundance profile of the selected bins confirmed the results obtained with the reference-based approach, returning similar differences in the prevalence of specific taxonomic units between cases and controls (Figure 4).

In detail, *Enterococcus faecalis* and *Enterococcus avium* species and the *Streptococcus* genus, belonging to the order *Lactobacillales*, predominated in the STEC positive samples (Figure 4A). Concerning the *Clostridiales* order, members of *Peptostreptococcaceae* and *Clostridiaceae* families showed a high abundance in the cases group (Figure 4A), while *Lachnospiraceae* and *Ruminococcaceae* spp. (*Ruminococcus gnavus*, *Faecalibacterium prausnitzii*) showed a high representation in the set of samples from healthy and recovered subjects, confirming their association with a healthy status in the human intestine (Figure 4B).

Finally, the approach of reconstructing single genomes from complex microbial communities confirmed a clear prevalence of members of the *Bifidobacteriales* order in the control group (Figure 4B).

For the analysis of the metagenomes from Crohn's disease, we selected nine bins based on a completeness $>35\%$, a contamination $<5\%$, and an abundance value $>2\%$ in at least one sample.

This analysis confirmed the lower complexity of the intestinal microbiota in Crohn's disease and the high representation of *E. coli* species in Sample 2 identified with the read mapping approach (Figure 3B).

DISCUSSION AND CONCLUSION

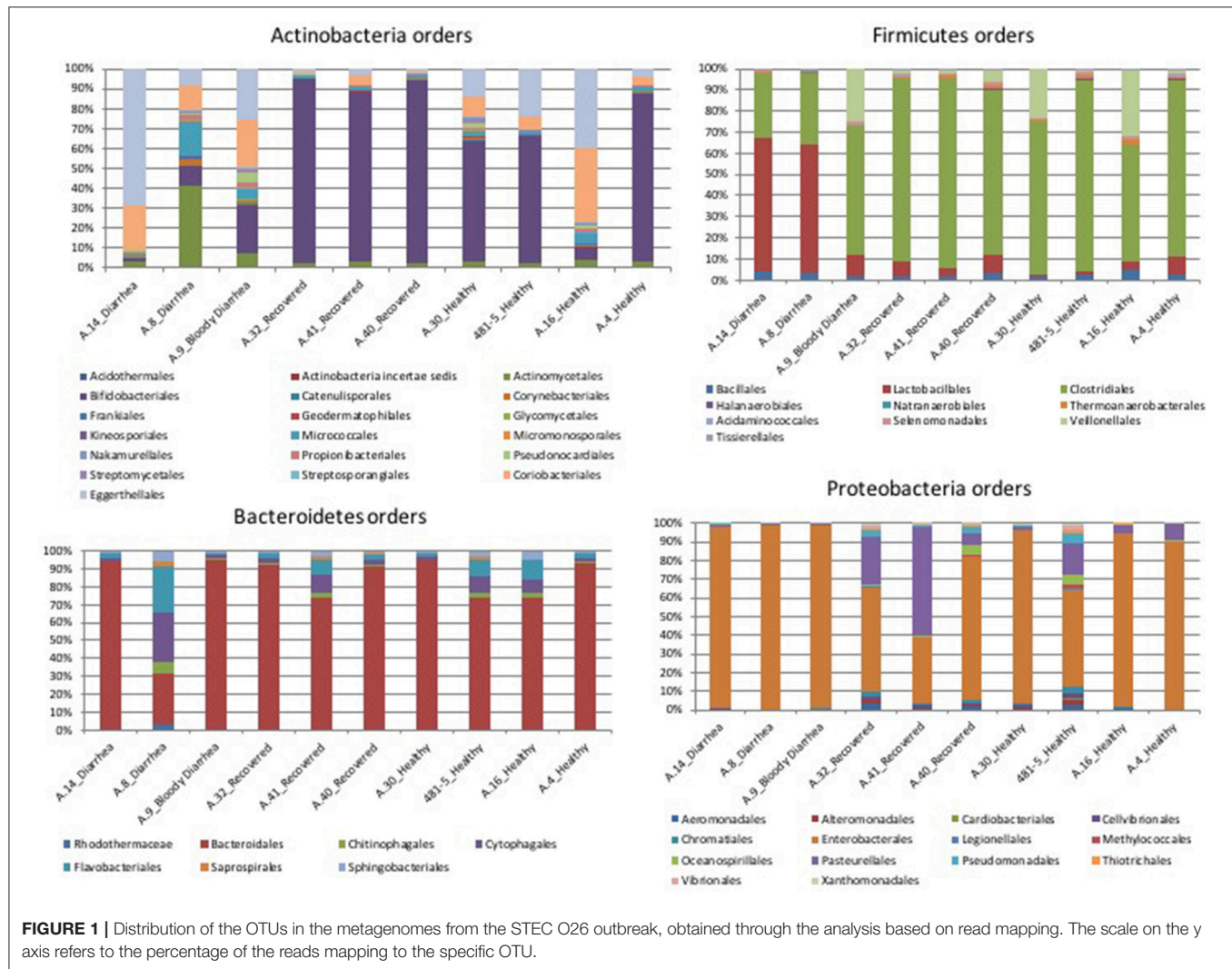
The human gastrointestinal tract microflora, comprise $\sim 10^{14}$ microbial cells that live in a mutual beneficial relationship with the host (Ley et al., 2006). Indeed, the gut microorganisms have a remarkable impact on human physiology, because they modulate the normal intestinal functions, produce vitamins and contribute to obtain energy from the food (Bäckhed et al., 2005). They have also a profound influence on the local and systemic immune responses and interfere with pathogen's colonization (Shreiner et al., 2015; Forbes et al., 2016).

In the present study, we aimed at using metagenomics as a tool to investigate possible changes in the composition of the intestinal microbiota in patients with STEC infection compared to healthy controls. Shotgun metagenomic sequencing was used to identify the presence of virulence genes associated with STEC strains and to perform a taxonomic classification of the microbial communities present in fecal samples collected from subjects

TABLE 3 | Relative abundance of the most abundant bacterial phyla in the samples from STEC-infected and healthy subjects, obtained through analysis based on read mapping.

| | A.9 (%) | A.8 (%) | A.14 (%) | A.40 (%) | A.41 (%) | A.32 (%) | A.4 (%) | A.16 (%) | 481-5 (%) | A.30 (%) |
|-----------------|---------|---------|----------|----------|----------|----------|---------|----------|-----------|----------|
| Actinobacteria | 0.5 | 1.1 | 6.3 | 56.2 | 14.8 | 50 | 14 | 3.5 | 28.5 | 0.8 |
| Proteobacteria | 15.7 | 56.6 | 13.3 | 1.9 | 3.4 | 1.6 | 29.5 | 12.2 | 1.6 | 3.7 |
| Bacteroidetes | 63.1 | 0.2 | 7.46 | 6.6 | 1 | 11 | 3.2 | 5 | 1.3 | 51 |
| Firmicutes | 19.7 | 41.3 | 71.2 | 34.2 | 79 | 36.7 | 52 | 76.4 | 66.5 | 16.7 |
| Verrucomicrobia | 0.16 | 0.0 | 0.0 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 | 0.3 | 27.5 |

Samples A.9, A.8, A.14 are from diseased patients; A.40, A.41, A.32 are from recovered patients; A.4, A.16, 481-5; A.30 are from healthy subjects.

**FIGURE 1 |** Distribution of the OTUs in the metagenomes from the STEC O26 outbreak, obtained through the analysis based on read mapping. The scale on the y axis refers to the percentage of the reads mapping to the specific OTU.

involved in an outbreak caused by a STEC O26:H11 strain as well as from patients with Crohn's disease, with and without evidence of STEC and other pathogenic *E. coli* infections.

The identification of all the virulence features of the epidemic STEC strain in the metagenomes, previously determined through whole genome sequencing of the isolated outbreak strain (unpublished), showed that metagenomics is a promising diagnostic tool for infectious diseases, subjected to the availability

of comprehensive databases of markers for the pathogens of interest. In sample A.14 the presence of the *iroN* gene, not present in the epidemic strain, was observed. The *iroN* gene encodes receptors involved in iron acquisition by bacteria (Russo et al., 2002), and it may have been present in a different co-circulating *E. coli* strain or even other bacteria.

The analyses carried out did not show the presence of genes encoding for Stx in specimens collected from both healed and

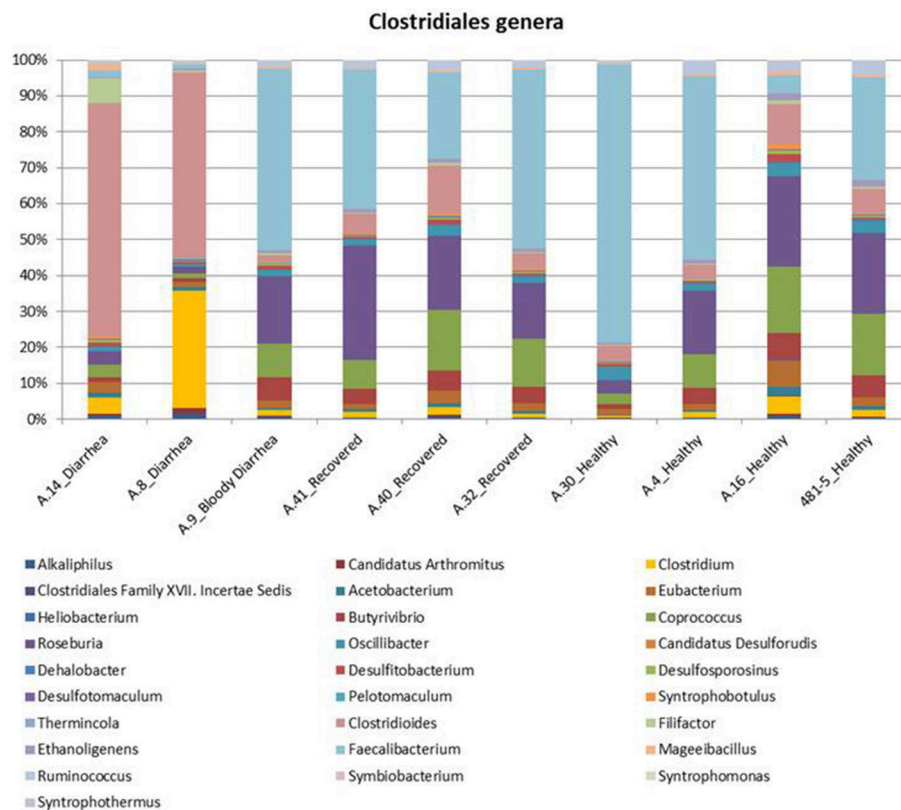
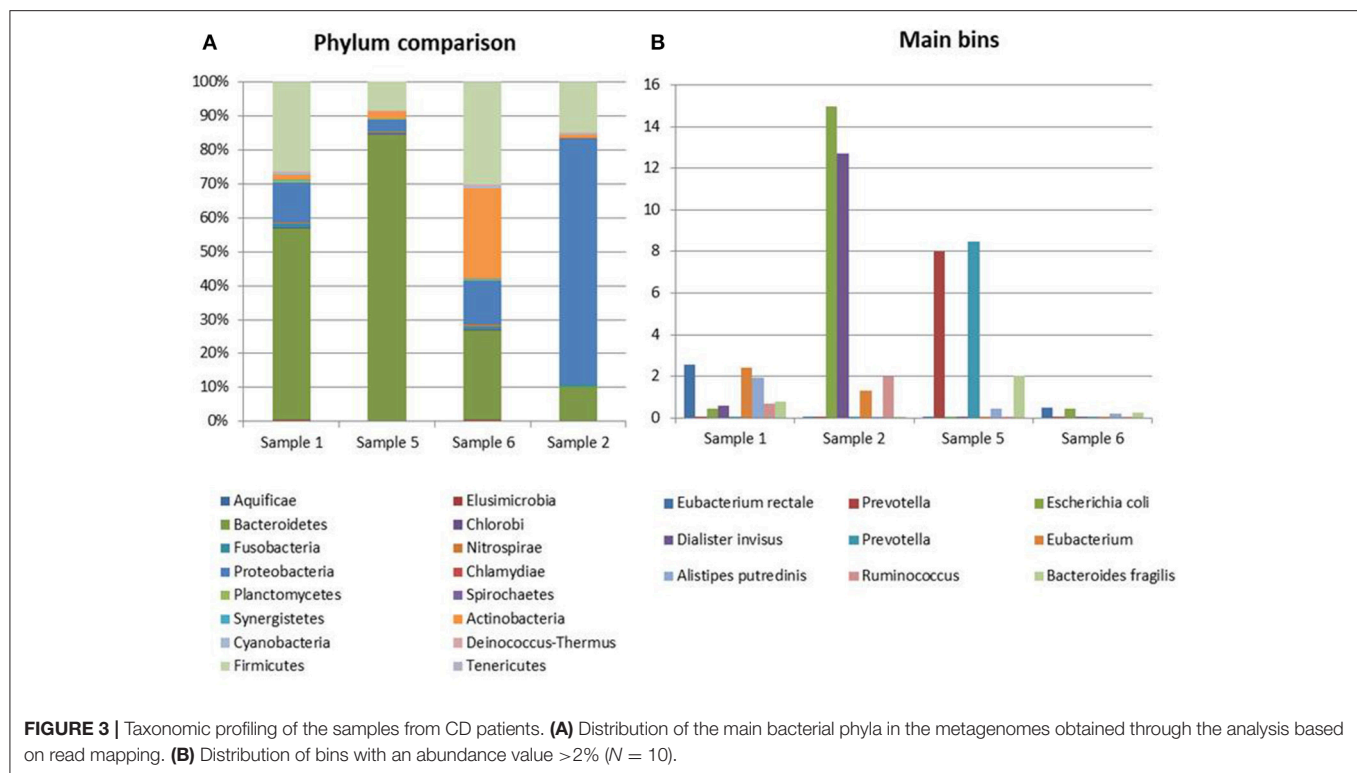


FIGURE 2 | Distribution of the main genera belonging to the *Clostridiales* order in the metagenomes from the STEC O26 outbreak, obtained through the analysis based on read mapping. The scale on the y axis refers to the percentage of the reads mapping to the specific OTU.

healthy subjects, indicating a good specificity of the approach used. Interestingly, in samples A.32, A.40, and A.41 (healed patients) evidence of a previous STEC infection was provided by the presence of some genes known to be located on the large virulence plasmid of STEC (**Table 1**). It is possible that this plasmid may have remained in the bacterial population after the STEC strain was cleared, leaving traces of the previous STEC infection.

The analysis of the metagenomes obtained from patients with Crohn's disease also confirmed previously obtained real-time PCR results (**Table 2**). For Sample 1 and Sample 5 this approach failed to identify the *escV* gene, associated with Enteropathogenic *E. coli* (EPEC). However, in these samples the evidence of EPEC was provided by the identification of the presence of other genes characteristic of this pathogroup, but different from those used in the real-time PCR experiments (**Table 2**). Indeed, while the real-time PCR approach may be more sensitive, due to the exponential amplification of the targets, the metagenomic approach may complement the lower technical sensitivity with the simultaneous search for more determinants associated with the same pathogroup. The use of a wider panel of targets may confer a stronger ability to detect pathogens, but it requires the availability of accurate and curated databases with the genetic markers characteristic for any of the pathogen of interest.

The taxonomic profiling of the STEC-infected samples provided insight into the changes occurring in the intestinal microbiota upon STEC infection. As previously described by other authors (Tap et al., 2009), our results showed that the *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Actinobacteria* were the bacterial phyla that prevailed in the samples, confirming their proposed role in maintaining the homeostasis in the human intestine (Tap et al., 2009) (**Table 3**). A deeper taxonomic classification highlighted a marked prevalence of some members of the *Bifidobacteriales* and *Clostridiales* orders in the STEC negative samples (**Figures 1, 2, 4B**). It has been already described that members of the *Bifidobacterium* genus confer positive health benefits to the human host (O'Callaghan and van Sinderen, 2016). These bacteria exert a probiotic action, stimulating the immune system (Perdigon et al., 1995), and providing a protection against gastrointestinal pathogens colonization by competitive exclusion of enteropathogens based on common binding sites on epithelial cells (Gueimonde et al., 2007). In addition, Fukuda and colleagues demonstrated that acetate production by *B. longum* strains is linked to the *in vitro* protection of host epithelial cells from the effect of Shiga toxin (Fukuda et al., 2011). The high abundance of *B. longum* and *B. breve* species, and *Bifidobacterium* genus observed in the samples collected from healthy and healed patients (**Figure 4B**)



could have different explanations. Such a high proportion of *Bifidobacterium* genus in the intestine may have been effective in protecting the healthy subjects against the STEC infection or may have favored the positive outcome of the infection in the healed patients, which have not developed the HUS. On the other hand, the high abundance observed of the members of this genus could have been the effect of a cleared or absent infection.

Similarly, the high abundance of the *Clostridiales* order in the STEC negative samples and the opposite prevalence of *Lactobacillales* order in the STEC positive samples could be put into relation with the STEC infection. As a matter of fact, it is possible that the intestinal colonization by STEC may contrast the normal permanence of *Clostridia* in the intestine, favoring the presence of *Bacilli*. It is interesting to note that *Faecalibacterium*, *Roseburia*, *Coprococcus*, and *Butyrivibrio* genera, highly prevalent in the metagenomes belonging to the control group, have all been proved to have a beneficial role in the human host intestine (Hugon et al., 2017). In this study, we have included some samples from patients with Crohn's disease. It has been described that patients with CD show a global decrease in the biodiversity of the fecal microbiota, essentially due to a markedly reduced diversity of *Firmicutes* (Manichanh et al., 2006; Sokol et al., 2008). As a matter of fact, a significantly reduced abundance of *Roseburia*, *Faecalibacterium*, and other genera belonging to the *Clostridiales* order in the intestine of CD patients has been reported (Kang et al., 2010; Morgan et al., 2012; Gevers et al., 2014). We considered the CD samples as showing taxonomic profiles related with a general status of dysbiosis not related with STEC infection and have evaluated

the differences observed between the samples from the cases and the controls collected in the framework of the STEC outbreak in the light of the profiles observed in the dysbiotic CD specimens. Our results confirmed the low biodiversity of the gut microbiota in CD patients and the lower proportion of members of the *Firmicutes* phylum (Figure 3 and Figure S1). Additionally, our findings highlighted the absence of beneficial *Bifidobacterium* species in the feces collected from CD subjects, similarly to what was observed in the samples from STEC infections (Figures 1, 4). This latter observation, confirms that a decrease in the abundance of *Bifidobacterium* is associated with an infection with diarrheagenic agents or a dysbiosis status as it has been previously proposed (Gevers et al., 2014; O'Callaghan and van Sinderen, 2016; He et al., 2017), suggesting that the perturbation in the proportions of the *Bifidobacterium* genus observed in this study may not be specific for the infection with STEC.

Our results indicate that metagenomics is effective in detecting genomic traits associated with STEC in stool samples from infected subjects, making it a promising tool for the culture-independent diagnosis of the infections. Additionally, the bioinformatic procedures used can be automated and applied to the detection of other pathogens. The analysis of the taxonomic composition of the intestinal microbiota showed a good agreement between the data obtained with both the reference-free and read mapping approaches, supporting the following comparative analyses. In this respect, to the best of our knowledge, our data provide a first evidence of the changes occurring in the intestinal microbiota of children in the course of

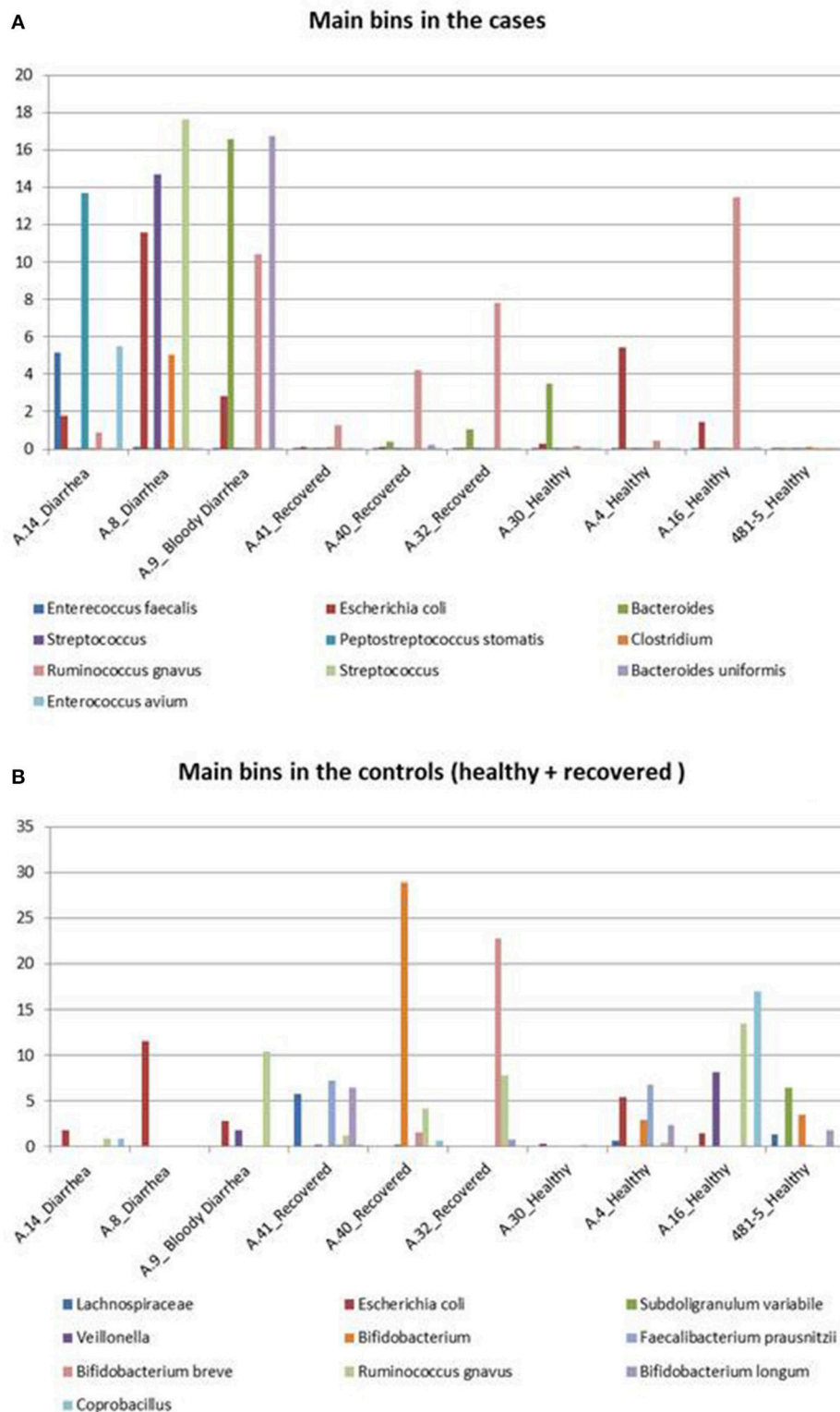


FIGURE 4 | Taxonomic profiling of the metagenomes from the STEC O26 outbreak using a reference-free approach. The figure highlights the bins identified in the cases and controls. **(A)** Shows the distribution of the 10 bins more abundant in the cases group across all the samples, while **(B)** illustrates the 10 bins which are more represented in the controls group, including healed and healthy subjects across all the samples.

STEC infection. Further studies are required to assess the reasons underlying such differences and if certain taxonomic profiles may be considered effective in protecting the host from acquiring the STEC infection.

AUTHOR CONTRIBUTIONS

FG performed the DNA extraction, the metagenomic sequencing, the bioinformatic analyses, and drafted the manuscript; FvM contributed to the bioinformatic analyses and revised the manuscript; AK installed the server for the bioinformatic analyses and provided assistance with the data analysis; VM contributed to the sequencing and bioinformatic analyses on the epidemic strain; GS contributed to the samples' selection and critically revised the manuscript; FM contributed to the isolation of the epidemic strain; BD contributed to the bioinformatic analyses and revised the manuscript; HA collected

the samples from Crohn's disease patients; GR contributed to the metagenomic sequencing; AF contributed to the metagenomic sequencing and revised the manuscript; JR contributed to the metagenomic sequencing and revised the manuscript; SM conceived the study and strongly contributed to revise the manuscript.

ACKNOWLEDGMENTS

BD was supported by the Netherlands Organization for Scientific Research (NWO) Vidi grant 864.14.004.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2018.00025/full#supplementary-material>

REFERENCES

- Armstrong, G. L., Hollingsworth, J., and Morris, J. G. (1996). Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiol. Rev.* 18, 29–51. doi: 10.1093/oxfordjournals.epirev.a017914
- Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920. doi: 10.1126/science.1104816
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Micheltore, R. W., Eisen, J. A., et al. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* 2:e415. doi: 10.7717/peerj.415
- Beutin, L., Bode, L., Ozel, M., and Stephan, R. (1990). Enterohemolysin production is associated with a temperate bacteriophage in *Escherichia coli* serogroup O26 strains. *J. Bacteriol.* 172, 6469–6475. doi: 10.1128/jb.172.11.6469-6475.1990
- Brunner, W., Schmidt, H., and Karch, H. (1997). EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol. Microbiol.* 24, 767–778. doi: 10.1046/j.1365-2958.1997.3871751.x
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Cambuy, D. D., Coutinho, F. H., and Dutilh, B. E. (2016). Contig annotation tool CAT robustly classifies assembled metagenomic contigs and long sequences. *bioRxiv:072868*. doi: 10.1101/072868
- Caprioli, A., Morabito, S., Brugère, H., and Oswald, E. (2005). Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet. Res.* 36, 289–311. doi: 10.1051/vetres:2005002
- de Vos, W. M., and de Vos, E. A. (2012). Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr. Rev.* 70(Suppl. 1), S45–S56. doi: 10.1111/j.1753-4887.2012.00505.x
- Forbes, J. D., Van Domselaar, G., and Bernstein, C. N. (2016). The gut microbiota in immune-mediated inflammatory diseases. *Front. Microbiol.* 7:1081. doi: 10.3389/fmicb.2016.01081
- Fujiwara, S., Hashiba, H., Hirota, T., and Forstner, J. F. (2001). Inhibition of the binding of enterotoxigenic *Escherichia coli* Pb176 to human intestinal epithelial cell line HCT-8 by an extracellular protein fraction containing BIF of *Bifidobacterium longum* SBT2928: suggestive evidence of blocking of the binding receptor ganglioside GM1 on the cell surface. *Int. J. Food Microbiol.* 67, 97–106. doi: 10.1016/S0168-1605(01)00432-9
- Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., et al. (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469, 543–547. doi: 10.1038/nature09646
- Garage, S. D., Strasser, J. E., Chalk, C. L., and Weiss, A. A. (2003). Nonpathogenic *Escherichia coli* can contribute to the production of Shiga toxin. *Infect. Immun.* 71, 3107–3115. doi: 10.1128/IAI.71.6.3107-3115.2003
- Garza, D. R., and Dutilh, B. E. (2015). From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cell. Mol. Life Sci.* 72, 4287–4308. doi: 10.1007/s00018-015-2004-1
- Gauthier, A., Puente, J. L., and Finlay, B. B. (2003). Secretin of the enteropathogenic *Escherichia coli* type III secretion system requires components of the type III apparatus for assembly and localization. *Infect. Immun.* 71, 3310–3319. doi: 10.1128/IAI.71.6.3310-3319.2003
- Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005
- Gotelli, N. J., and Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* 4, 379–391. doi: 10.1046/j.1461-0248.2001.00230.x
- Gueimonde, M., Margolles, A., de los Reyes-Gavilán, C. G., and Salminen, S. (2007). Competitive exclusion of enteropathogens from human intestinal mucus by Bifidobacterium strains with acquired resistance to bile—a preliminary study. *Int. J. Food Microbiol.* 113, 228–232. doi: 10.1016/j.ijfoodmicro.2006.05.017
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- He, Q., Gao, Y., Jie, Z., Yu, X., Laursen, J. M., Xiao, L., et al. (2017). Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience* 6, 1–11. doi: 10.1093/gigascience/gix050
- Hugon, P., Lagier, J. C., Colson, P., Bittar, F., and Raoult, D. (2017). Repertoire of human gut microbes. *Microb. Pathog.* 106, 103–112. doi: 10.1016/j.micpath.2016.06.020
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107
- Joensen, K. G., Scheut, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13

- Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. doi: 10.7717/peerj.1165
- Kang, S., Denman, S. E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., et al. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm. Bowel Dis.* 16, 2034–2042. doi: 10.1002/ibd.21319
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848. doi: 10.1016/j.cell.2006.02.017
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55, 205–211. doi: 10.1136/gut.2005.073817
- Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13:r79. doi: 10.1186/gb-2012-13-9-r79
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116
- O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W., and Formal, S. B. (1984). Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* 226, 694–696. doi: 10.1126/science.6387911
- O'Callaghan, A., and van Sinderen, D. (2016). Bifidobacteria and their role as members of the human gut microbiota. *Front. Microbiol.* 7:925. doi: 10.3389/fmicb.2016.00925
- Okuda, T., Tokuda, N., Numata, S., Ito, M., Ohta, M., Kawamura, K., et al. (2006). Targeted disruption of Gb3/CD77 synthase gene resulted in the complete deletion of globo-series glycosphingolipids and loss of sensitivity to verotoxins. *J. Biol. Chem.* 281, 10230–10235. doi: 10.1074/jbc.M600057200
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Perdigon, G., Alvarez, S., Rachid, M., Agüero, G., and Gobbato, N. (1995). Immune system stimulation by probiotics. *J. Dairy Sci.* 78, 1597–1606. doi: 10.3168/jds.S0022-0302(95)76784-4
- Rivière, A., Selak, M., Lantin, D., Leroy, F., and De Vuyst, L. (2016). Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Front. Microbiol.* 7:979. doi: 10.3389/fmicb.2016.00979
- Russo, T. A., McFadden, C. D., Carlino-MacDonald, U. B., Beanan, J. M., Barnard, T. J., and Johnson, J. R. (2002). IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*. *Infect. Immun.* 70, 7156–7160. doi: 10.1128/IAI.70.12.7156-7160.2002
- Shreiner, A. B., Kao, J. Y., and Young, V. B. (2015). The gut microbiome in health and in disease. *Curr. Opin. Gastroenterol.* 31, 69–75. doi: 10.1097/MOG.0000000000000139
- Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J. J., et al. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16731–16736. doi: 10.1073/pnas.0804812105
- Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J. P., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–2584. doi: 10.1111/j.1462-2920.2009.01982.x
- Tatsuno, I., Kimura, H., Okutani, A., Kanamaru, K., Abe, H., Nagai, S., et al. (2000). Isolation and characterization of mini-Tn5Km2 insertion mutants of enterohemorrhagic *Escherichia coli* O157:H7 deficient in adherence to Caco-2 cells. *Infect. Immun.* 68, 5943–5952. doi: 10.1128/IAI.68.10.5943-5952.2000

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gigliucci, von Meijenfildt, Knijn, Michelacci, Scavia, Minelli, Dutilh, Ahmad, Raangs, Friedrich, Rossen and Morabito. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.